Structure of uncomplexed and linoleate-bound *Candida cylindracea* cholesterol esterase

Debashis Ghosh[1,2]*, Zdzislaw Wawrzak[3†], Vladimir Z Pletnev[4], Naiyin Li[1], Rudolf Kaiser[5], Walter Pangborn[1], Hans Jörnvall[5], Mary Erman[1] and William L Duax[1]

[1]Hauptman-Woodward Medical Research Institute, Inc. (formerly the Medical Foundation of Buffalo, Inc.), 73 High Street, Buffalo, NY 14203-1196, USA, [2]Roswell Park Cancer Institute, Buffalo, NY 14263, USA, [3]Technical University of Lodz, Institute of Physics, ul. Wolczanska 219, 93-005 Lodz, Poland, [4]Shemyakin Institute of Bioorganic Chemistry, Russian Academy of Sciences, ul. Miklukho-Maklaya, 117871 GSP-7 Moscow V-427, Russia and [5]Department of Medical Biochemistry and Biophysics, Karolinska Intitutet, S-171 77 Stockholm, Sweden

Background: *Candida cylindracea* cholesterol esterase (CE) reversibly hydrolyzes cholesteryl linoleate and oleate. CE belongs to the same α/β hydrolase superfamily as triacylglycerol acyl hydrolases and cholinesterases. Other members of the family that have been studied by X-ray crystallography include *Torpedo californica* acetylcholinesterase, *Geotrichum candidum* lipase and *Candida rugosa* lipase. CE is homologous to *C. rugosa* lipase 1, a triacylglycerol acyl hydrolase, with which it shares 89% sequence identity. The present study explores the details of dimer formation of CE and the basis for its substrate specificity.

Results: The structures of uncomplexed and linoleate-bound CE determined at 1.9 Å and 2.0 Å resolution, respectively, reveal a dimeric association of monomers in which two active-site gorges face each other, shielding hydrophobic surfaces from the aqueous environment. The fatty-acid chain is buried in a deep hydrophobic pocket near the active site. The positioning of the cholesteryl moiety of the substrate is equivocal, but could be modeled in the hydrophobic core of the dimer interface.

Conclusions: The monomer structure is the same in both the complexed and uncomplexed crystal forms. The dimers differ in the relative positions of the two monomers at the dimer interface. Of the 55 residues that are different in CE from those in *C. rugosa* lipase 1, 23 are located in the active site and at the dimer interface. The altered substrate specificity is a direct consequence of these substitutions.

Structure 15 March 1995, **3**:279–288

Key words: cholesterol esterase, cholesteryl linoleate, dimer formation, hydrolysis and esterification, X-ray structure

Introduction

Cholesteryl linoleate and oleate are the major components of arterial plaque [1]. Mammalian cholesterol esterase (CE) enzymes in the intestine and pancreas are responsible for the adsorption of free cholesterol through hydrolysis of dietary fat and re-esterification of cholesteryl esters. Although cholesterol esterases found in fungi are probably involved in the utilization of fat as a source of energy, they may be useful models for studying the mechanism of action of their mammalian counterparts. CE from the fungus *Candida cylindracea* is a glycoprotein that belongs to the same lipase/esterase family as *Candida rugosa* lipase 1 (CRL) [2], *Torpedo californica* acetylcholinesterase (AChE) [3] and *Geotrichum candidum* lipase (GCL) [4], all of which contain Ser-Glu-His as the catalytic triad and comprise ~540 residues. While AChE is a cholinesterase, CRL and GCL are primarily triacylglycerol acyl hydrolases. Cholesterol esters of long-chain fatty acids are not reported to be substrates for these enzymes. CE, on the other hand, specifically accepts cholesteryl linoleate and oleate as substrates [5]. Cygler *et al.* [6] have studied the sequence homology of 32 members of this family that vary in length within the range 496–592 amino acids. Although highly homologous, the location and nature of the sequence differences between CE and CRL are consistent with altered substrate specificity. CE shares 40–50% sequence identity with GCL and AChE. GCL, AChE, CRL and other members of the α/β hydrolase superfamily [7,8], namely *Rhizomucor miehei* lipase [9,10], human pancreatic lipase [11], a porcine procolipase+human pancreatic lipase complex [12], *Fusarium solani* cutinase [13], and *Candida antartica* lipase B [14] have been studied by X-ray crystallography. Together with these apo-enzymes, crystallographic studies of complexes of *R. miehei* lipase with an inhibitor [15] and of lipase–procolipase with phosphatidylcholine [16] provide insight into the mechanism of action of this group of enzymes. In particular, the structures of GCL, AChE and CRL are the basis for initial understanding of the complex biochemical machinery responsible for hydrolysis of lipids at a water–lipid interface.

We report the crystal structure determination of CE in two triclinic crystal forms, the apo-enzyme (apo-CE), and the enzyme–linoleate complex (lino-CE) which was grown in the presence of cholesterol linoleate. These crystal forms, apo-CE and lino-CE, have been referred to as Type II and Type III, respectively, in

*Corresponding author. †Present address: Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, PA 19104-6323, USA.

© Current Biology Ltd ISSN 0969-2126

preliminary publications [17,18]. Closely associated homodimers are found in the unit cells of both crystal forms. The existence of a dimer of CE in solution has also been shown [18] and there is biochemical evidence for the existence of dimeric forms of porcine cholesterol esterase [19,20].

Tertiary structure

The conformations of the four CE monomers observed in the two crystal structures reported here are essentially the same. The overall fold of CE can be described as a single-domain α/β protein with a mixed 13-stranded β-sheet at the core surrounded by 16 α-helices (Figs 1a and 1b). A major segment of the sheet comprises seven parallel strands, β6–β12. The active site contains a Ser209-Glu341-His449 catalytic triad. Ser209 adopts an unfavorable conformation, as has been observed in many peptidases, esterases and hydrolases [21]. The subunit tertiary structure of CE is similar to that of AChE and GCL, except in some loops and turns and the 'flap' region (residues 65–91, a region that usually sits on the active-site cavity in other lipases, e.g. in GCL [4]). There is also variation in the lengths of some of the secondary-structural elements. When the Cα-atoms of the catalytic triad residues were superimposed, root mean square deviations (rmsds) of 0.40 Å, 0.25 Å and 0.18 Å were calculated for the pairs CE/AChE, CE/GCL, and CE/CRL, respectively. The structures of CE and CRL are homologous, reflecting a very high degree of sequence identity. In both apo-CE and lino-CE, association of monomers is observed and dimers are formed in which the active-site gorges from the two subunits face each other creating a cavity lined with hydrophobic and aromatic residues. Very similar dimer formation occurs in the crystal packing between two crystallographic two-fold related monomers in CRL [2]. Nonetheless, there are some important substitutions near the catalytic triad, in the active-site gorge and at the dimer interface that are probably responsible for the altered substrate specificity, in addition to causing local differences in the structure at the areas of substitution.

In Fig. 2, dimers of apo-CE (orange), lino-CE (green) and CRL (blue) are superimposed by least-squares fitting of the Cα-atoms of one of the monomers. At the dimer interface of CE a close interaction is observed between identical segments of polypeptide chains related by the non-crystallographic two-fold axis including α-helical residues 85–90 at the top of the interfacial region, and segments 396–400 at the bottom (away from the viewer in Fig. 2). The two antiparallel segments form four interchain hydrogen bonds that are between 2.99 Å and 3.38 Å long. There are two openings at the interface and two more generated by the non-crystallographic two-fold axis. The central interfacial cavity contains two active sites. The opening shown in Fig. 3a is bordered by helical residues 73–86 from the A-subunit and residues 89–91, 130–134 and 453–460 from the B-subunit. The second opening, shown in Fig. 3b, is directly across the cavity from the first, and is lined by residues 439–445, 450–455, 460–462 and 506–509 from the A-subunit, and residues 296–301, 66–72 and 344–353 from the B-subunit. No solvent molecules could be located at the dimer interface between protein atoms from the two subunits, although solvent water molecules are found in the catalytic cavity, near Ser209. The only other close approach between subunits A and B, in apo-CE or lino-CE, is an isolated contact between the main-chain carbonyl of

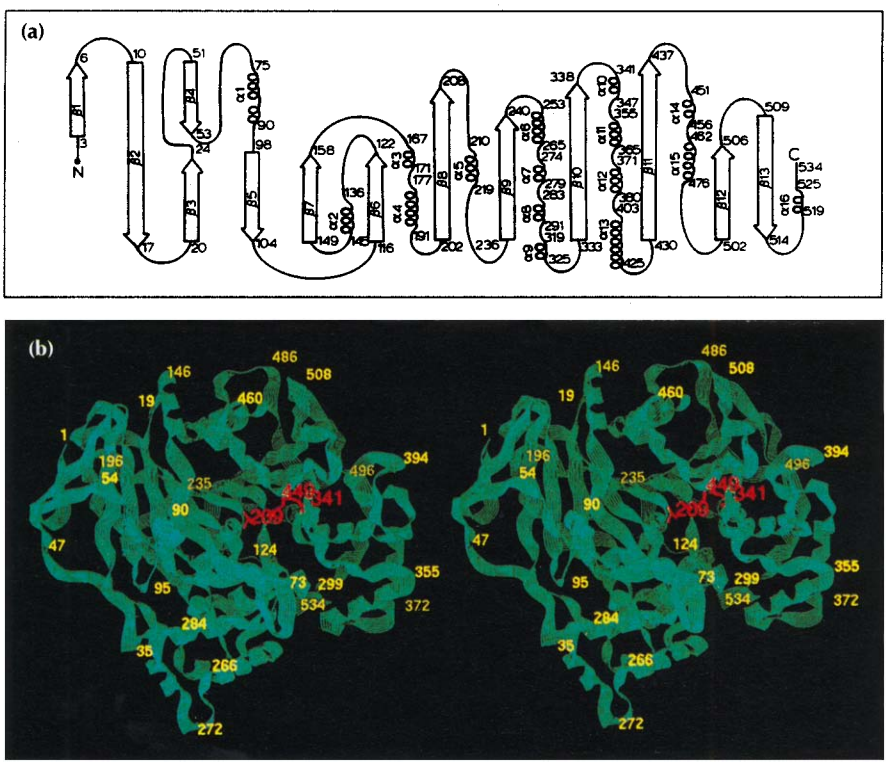


Fig. 1. **(a)** Schematic diagram of the structure of CE showing the overall fold, distribution of the secondary-structure elements and their relative locations. Arrows indicate β-strands and coils α-helices. The beginning and end residue numbers of β-strands and α-helices are indicated. **(b)** Ribbon diagram of the A-subunit showing the course of the main chain. (Figure drawn by CHAIN [29].)

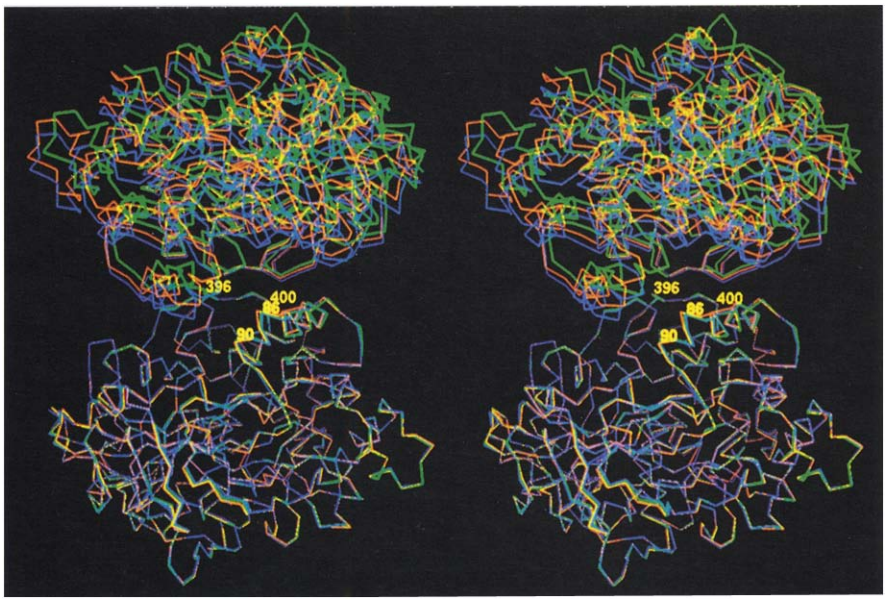


Fig. 2. Stereoview of the superposition of Cα-chains of apo-CE (orange), lino-CE (green) and CRL (blue) dimers. The monomers of apo-CE and CRL have each been fitted to lino-CE by a least-squares procedure. The lower half of the drawing illustrates the excellent agreement in the fit of the Cα-chains. The upper half of the drawing illustrates the variation in the relative positions of the monomers in the dimers observed in the three crystals. The closest dimer association between atoms from two subunits in the same structure occurs in CRL (blue) and the most displaced in lino-CE (green). (Figure drawn by CHAIN [29].)

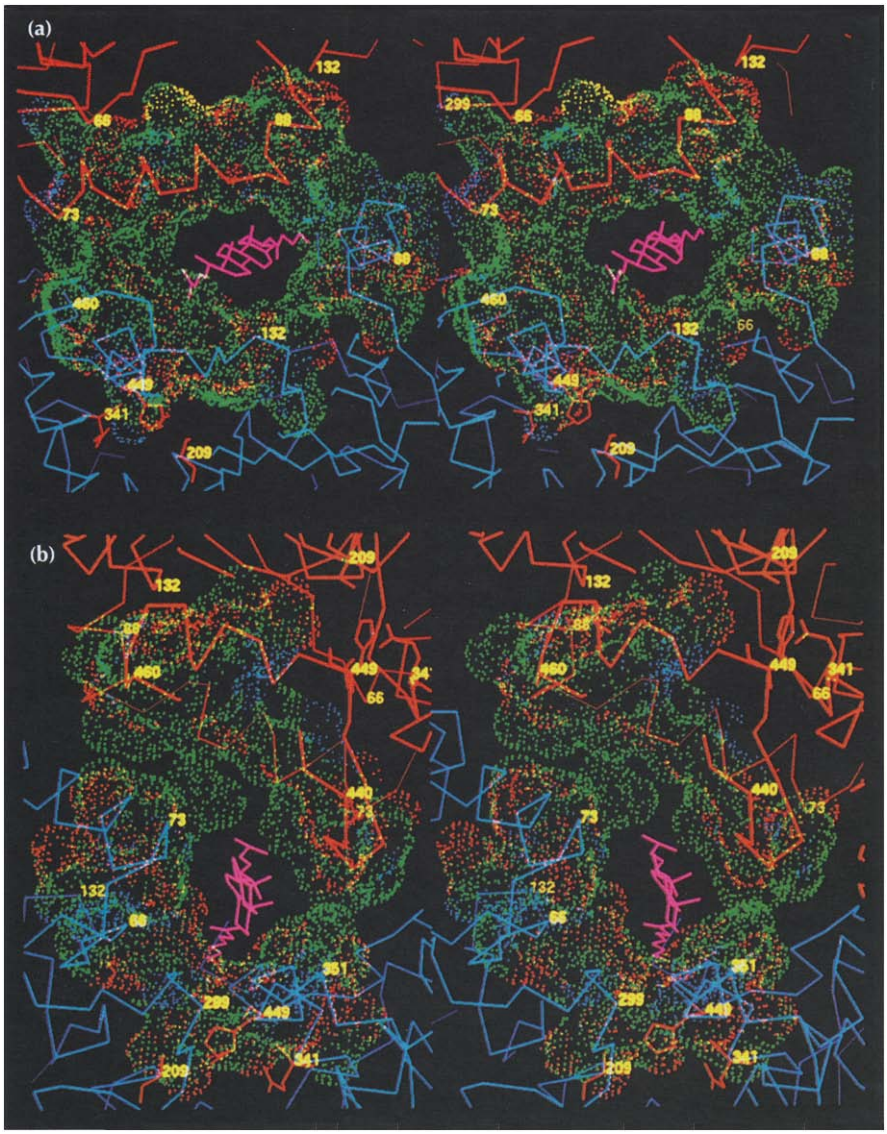


Fig. 3. Stereoviews of two openings to the active site at the lino-CE dimer interface. The Cα backbones of the A- and B-subunits are shown in red and blue, respectively. A color-coded Connolly surface (green, non-polar; red, polar and yellow, sulfur) is shown around residues that border the interfacial openings. **(a)** The first opening. **(b)** The second opening, directly across from the first. In each view, a model of a cholesteryl ester molecule shown in pink is positioned with its widest section at the opening, illustrating that the interfacial openings are sufficiently large to permit access of the substrate to the active site. Side chains of the catalytic triad (Ser209-His449-Glu341) are shown in red. For detailed descriptions of interfacial residues bordering the two openings, see the text. (Figures drawn by CHAIN [29].)

Gly230 of the A-subunit with the main-chain amide of Thr430 of the B-subunit.

Although the subunit structures in apo-CE and lino-CE are the same, the subunit association is slightly different. In Fig. 2, the 534 Cα-atoms of the A-subunits of the two structures have been least-squares fitted and have an rmsd of 0.23 Å. This fit gives an rmsd for the Cα-atoms in the B-subunits of 3.68 Å. When the A-subunits are least-squares fitted, the atoms of the B-subunits with the smallest deviations between the two sets of Cα-atoms (rmsds in the range 0.40–1.00 Å) are those nearest to the segment 396–400. Very similar values are obtained when the B-subunits are least-squares fitted. In the lino-CE structure (green in Fig. 2), one subunit undergoes a rotation of ~5° about a point near the dimer interface where the interfacial segment 396–400 forms four hydrogen bonds with the same region of the other subunit. The rotation opens the other end of the dimer interface slightly. The rmsds between lino-CE subunits and two CRL monomers (in blue in Fig. 2) are 0.29 Å and 5.58 Å, for the fitted and unfitted subunits, respectively. The corresponding rmsds between apo-CE (orange in Fig. 2) and CRL are 0.32 Å and 2.56 Å, respectively.

For apo-CE, subunits A and B have accessible surface areas (using a probe of radius 1.5 Å) of 17310 Å^2 and 17289 Å^2, respectively. The apo-CE dimer has a total accessible surface area of 32736 Å^2. Therefore, 1863 Å^2 or 5.4% of the total monomer surface was inaccessible due to formation of the dimer. This value of ~900 Å^2 per monomer of surface area corresponds well to values observed for protein–protein interactions in antibody–antigen complexes [22] and to the increased hydrophobic area exposed upon activation of lipases [23]. The accessible surface area for the lino-CE dimer is 32944 Å^2 and the surface inaccessible due to dimerization is 1618 Å^2 or 4.7%. Thus, the apo-CE to lino-CE transition results in a 250 Å^2 increase in the interfacial surface. The calculation of the total accessible dimer surface does not exclude the ~4000 Å^2 of inner surface area, accessible through the openings described above.

Fig. 4 is a plot of the percentage of the total dimer surface made inaccessible upon dimerization, as a function of the sequence number. This plot was obtained by subtracting the accessible surface of each residue of the dimer from that of the monomer. The inaccessible part of the dimer interface is formed by residues 66–91, 344–353, 396–400 and 439–462, for both apo-CE and lino-CE structures. The apo-CE structure is more tightly associated than the lino-CE structure in the regions 66–91 and 453–462, which form significant portions of the two openings described above. There are no significant differences between apo-CE and lino-CE in the regions 344–353 and 396–400, the latter being the area of inter-subunit hydrogen bond formation.

The central interfacial cavity results from confluence of two hydrophobic gorges. This nearly spherical cavity, at

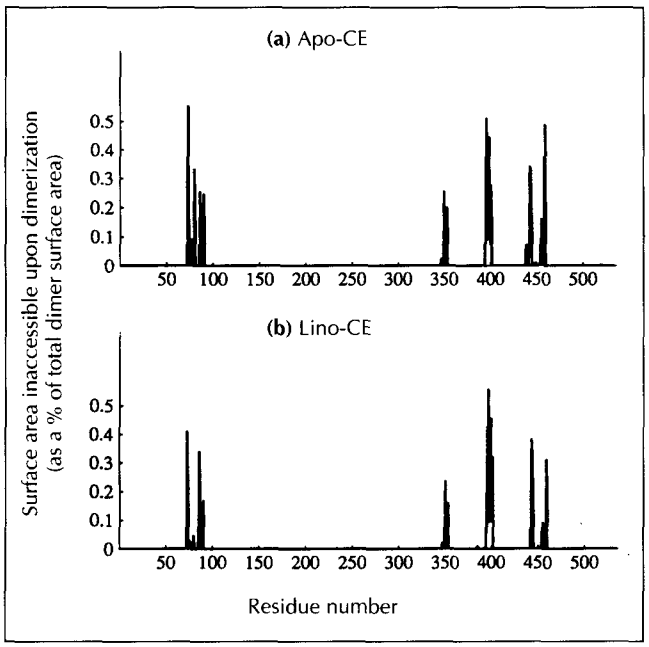
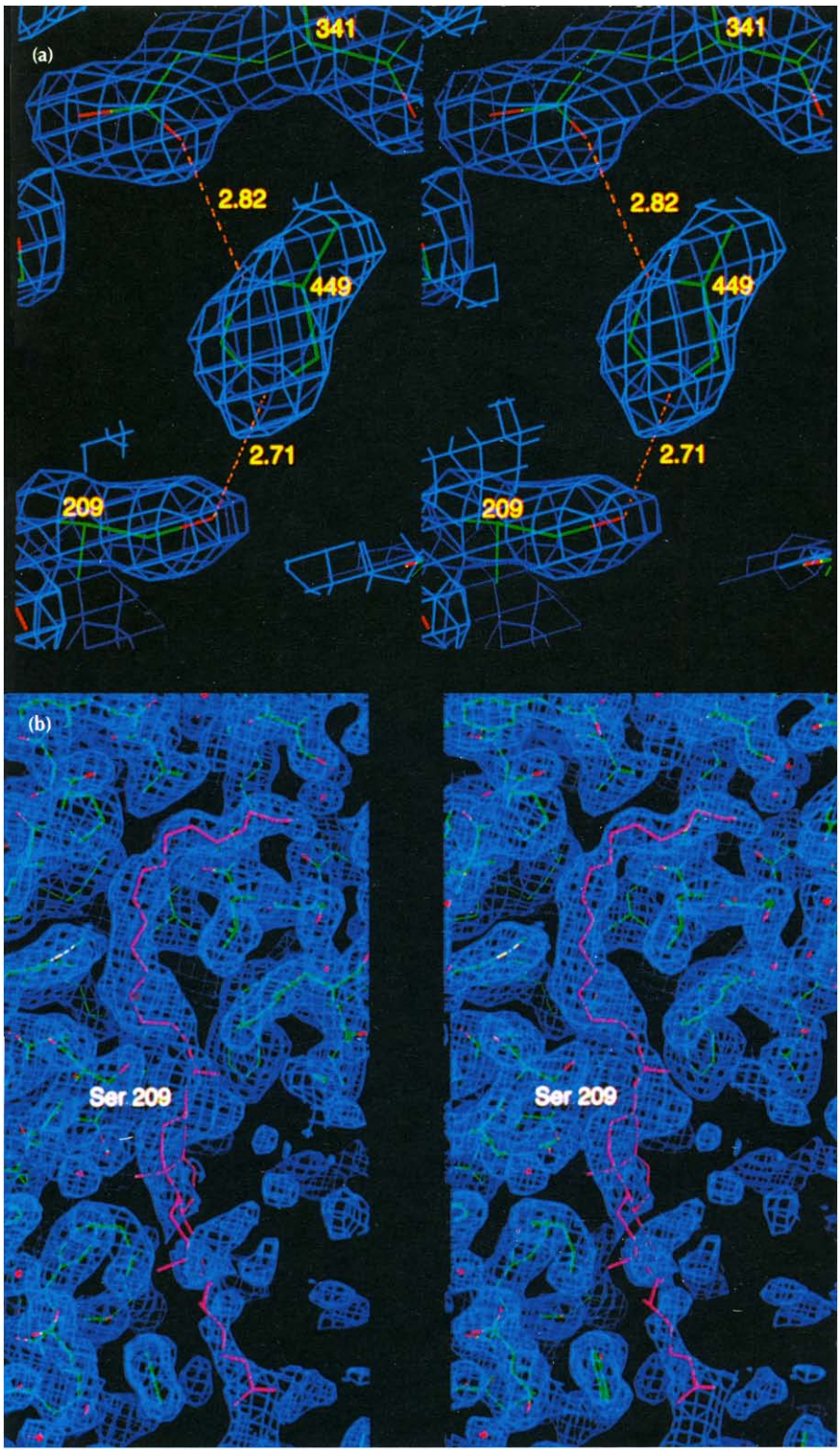


Fig. 4. Plots of the surface area that becomes inaccessible upon dimerization (as a percentage of the total dimer surface area) *versus* residue number. **(a)** Apo-CE structure. **(b)** Lino-CE structure. The peaks in the plots correspond to regions of close monomer–monomer association in the dimer.

the center of the dimer, is lined with hydrophobic and aromatic residues from several regions of each subunit, as follows: loop 62–67, helix 73–91, loop 122–133, turn 208–209, loop 296–301, helix 344–350 and part of helix 439–459. Extensive overlap occurs between the residues that are located at the interfacial region and the ones that make up the central hydrophobic core. Of the 55 residues that differ between CRL and CE, 23 are either at the active site, at the interface or within the hydrophobic core. Furthermore, in all but two cases (F344L and L417H), the variant residues found in CE are more hydrophobic than the corresponding residues in CRL. With substrate specificity altered from triacylglycerol to cholesteryl ester, it is not surprising that CE has a more hydrophobic substrate-binding pocket.

R. miehei lipase has been observed in significantly different conformations in different crystal forms. In one crystal form, one domain of the structure forms a lid or flap over the hydrophobic active-site gorge. The lid is lifted off of the gorge in a second crystal form. This observed flexibility provided the basis for a postulated mechanism of action of the enzyme at an oil–water interface [23]. The analogous flap of CE is open, exposing the hydrophobic gorge in all crystallographic studies reported thus far. In a polar environment, dimer formation might well constitute a mechanism by which the otherwise exposed hydrophobic core of the subunits could be sequestered.

Two of the three loops containing residues 62–92, 122–129 and 294–305, previously proposed to be involved in the catalytic mechanism in CRL [2], have major roles in the constitution of dimeric CE, either by



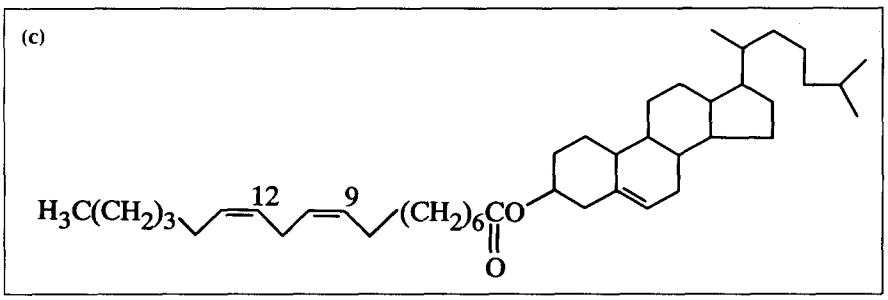


Fig. 5. (a) The catalytic triad, Ser209-His449-Glu341, as it is observed in the final ($3F_{obs}-2F_{calc}$) electron-density map contoured at 1.5σ. Hydrogen-bond formation among the residues is indicated. **(b)** A ($3F_{obs}-2F_{calc}$) map of the complex structure (lino-CE) near the active site of the A-subunit prior to the inclusion of the linoleate in the model, showing unbiased electron density for the refined cholesteryl linoleate molecule (in magenta). Ser209 of the catalytic triad is also shown. The map is contoured at 1.1σ. Surrounding protein and solvent atoms are shown in standard colors (carbon: green, nitrogen: blue, oxygen: red and sulfur: yellow). (Figures drawn by CHAIN [29].) **(c)** The chemical structure of cholesteryl linoleate.

forming interfacial interactions or by contributing to substrate specificity at the active site. It is possible that a functional role for the dimer is peculiar to cholesterol esterase. The four openings large enough to accommodate cholesterol and linoleate molecules (Fig. 3) may modulate substrate access to and product departure from the catalytic site. The subtle but real difference in association of apo-CE and lino-CE monomers may be related to substrate binding.

Binding of linoleate at the active site

The final model of the catalytic triad, Ser209-His449-Glu341, is shown in the ($3F_{obs}$–$2F_{calc}$) density map illustrated in Fig. 5a. Fig. 5b is a stereoview of the refined cholesteryl linoleate molecule in the unbiased ($3F_o$–$2F_c$) electron-density map in the A-subunit. The chemical substrate of cholesteryl linoleate is shown in Fig. 5c. The linoleate chain, including its *cis* bonds, could be readily positioned in the electron-density map within the catalytic cavity, and on one side of the catalytic triad (upper half of Fig. 5b). Only weak and discontinuous density was observed on the other side of the catalytic triad (lower half of Fig. 5b) and corresponds to the cholesteryl end of the substrate. The refined occupancy factor of 0.60 and 0.46 in the A- and B-subunits and the higher B-value of the substrate (~40 Å^2 compared with 20 Å^2 for the protein) suggest that a mixture of hydrolyzed fatty acids and intact cholesteryl esters has been captured in the complex, or that the cholesteryl end of the substrate is dynamically disordered. A combination of both these situations cannot be ruled out. The fatty-acid chain is imbedded in a deep hydrophobic cleft of the protein that traverses it from the internal catalytic cleft to the C terminus, near the outer surface of the protein. The entire length of the fatty-acid chain in lino-CE is surrounded by seven protein segments that are entirely hydrophobic: residues 121–124 at the C-terminal end of β6, residues 209–213 from α5, residues 296–304 from the loop between α8 and α9, residues 245 and 246 from the loop between β9 and α6, residues 410–415 of α13, residues 359–362 of α11, and C-terminal residues 532–534 (Fig. 6). The ester bond is in close proximity to the Ser209 side chain hydroxyl oxygen. The distance between the Oγ atom of Ser209 and the carbonyl carbon atom of the ester bond that becomes covalently linked to it in the tetrahedral intermediate [25] is 4.1 Å. The tetrahedral intermediate formed during catalysis could be stabilized by the main-chain amide groups of Gly123, Gly124 and Ala210. In recently published structures of CRL–inhibitor complexes, similar binding of fatty-acyl chains has been reported [24].

The end of the fatty-acid chain interacts with the C-terminal hydrophobic residues Phe532, Phe533 and Val534 (Fig. 6). The C terminus is tucked into a hydrophobic pocket surrounded by α11, α8 and residues 301–310. The free carboxylic end of the polypeptide forms a salt bridge with Arg309 and a hydrogen bond with the main-chain amide of Asp248. This suggests that the C terminus could also serve as a gate for release of the fatty-acyl product. When the C-terminal gate is open, Arg309 could neutralize the charge due to nearby Asp248 on the outer surface of the molecule. However, for the product to be released through this narrow hydrophobic aperture, the acyl group of the fatty acid would have to be protonated.

The cholesterol part of the substrate molecule that has been refined in the site is surrounded by residues 127–133 on one side, and residues 448–459 from α14 on the other. The cholesterol moiety is situated outside the catalytic gorge and has few close contacts with side-chain atoms (Fig. 6). The hydrophobic C(17) side chain of the cholesterol molecule, as modeled, extends across the dimer interface and approaches residues Leu73 and Ala77 of α1 of the other monomer. Thus, the cavity in the dimer is large enough to accommodate a complete cholesteryl linoleate molecule.

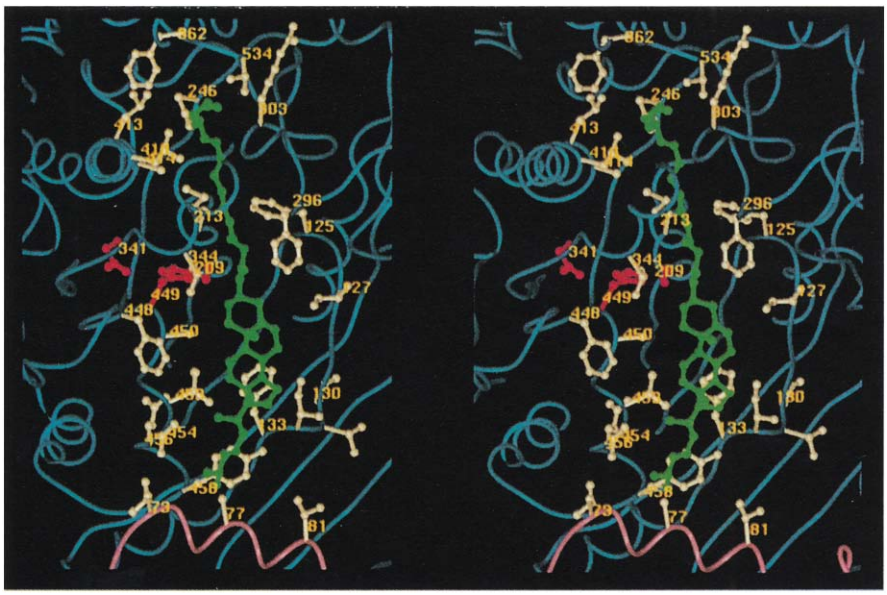


Fig. 6. Stereoview of the active site of the A-subunit of the lino-CE structure. The cholesteryl linoleate molecule is shown in green, the side chains of the catalytic triad (Ser209-Glu341-His449) in red, other side chains in the immediate neighborhood of the substrate in pale yellow, and the protein backbone in blue. The distance between the Ser209 hydroxyl oxygen and the carbonyl carbon of the ester bond of the substrate is 4.1 Å. The C-terminal residue, Val534, is also shown at the top edge of the figure. Part of the backbone of the B-subunit is visible (in pink) at the bottom of the figure. (Figure prepared using SETOR [33].)

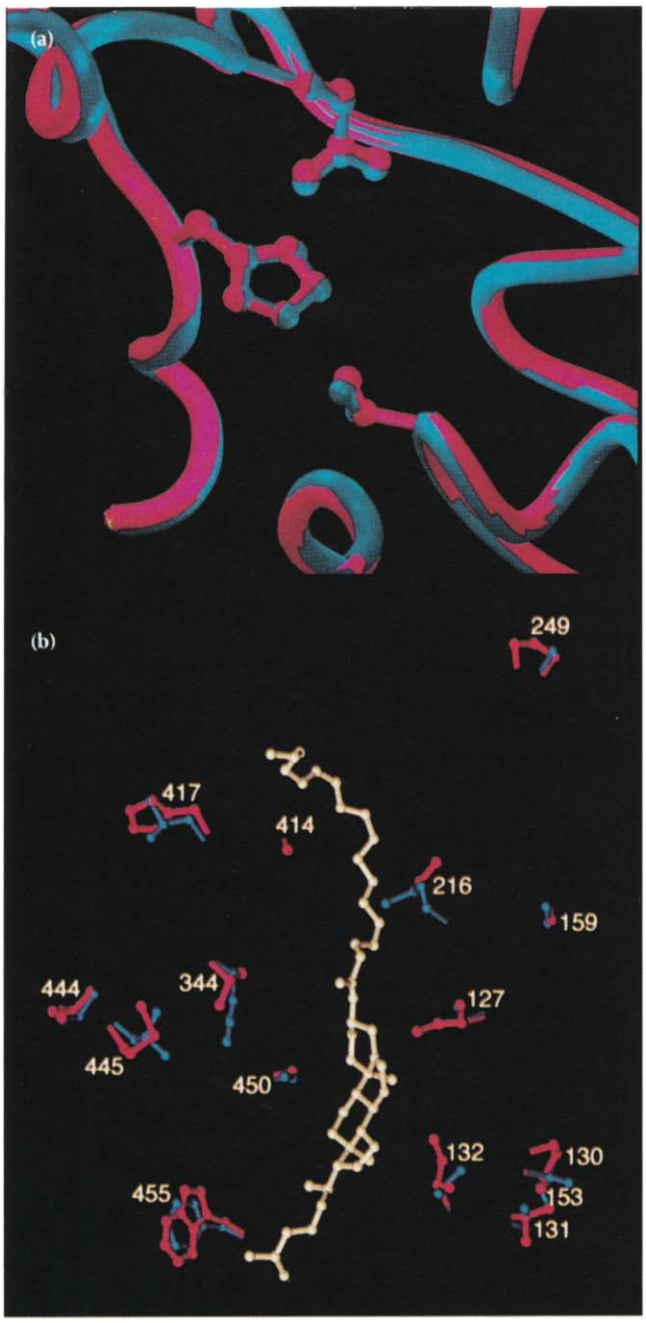


Fig. 7. (a) Superimposed active sites of lino-CE (red) and CRL (blue), showing the catalytic triads Ser209-His449-Glu341. **(b)** A view of the substrate cholesteryl linoleate (pale yellow) and the proximal side chains that differ between lino-CE and CRL. Side chains belonging to lino-CE are shown in red. (Figures prepared using SETOR [33].)

It is possible that during crystallization the substrate and/or cleavage products became trapped in the dimer. A local increase in product concentration or a significant drop in pH or hydration state in the catalytic site during crystallization could disrupt the catalytic mechanism. In apo-CE crystals, which were crystallized without the addition of cholesteryl linoleate, the substrate-binding pocket contains discontinuous residual density which may represent trapped solvent molecules.

Comparison with CRL: key amino acid differences

Fig. 7a shows a close-up view of the superimposed catalytic triads of lino-CE (red) and CRL (blue). The rmsd between the three Cα-atoms of the catalytic triads of CE and CRL is 0.18 Å. When all Cα-atoms of lino-CE and CRL are least-squares fitted the rmsd (0.29 Å) is only marginally higher than estimated random positional error and similar to the rmsd between the apo-CE and lino-CE structures. Regions of higher rmsds occur between these two structures which may be significant. One such region is the helix α1 of the open flap where the deviations are in the range 0.46–0.81 Å. Residues in the regions 128–140, 302–303, 363–368, 485–486 and 522–529 of CRL have deviations from the corresponding Cαs of lino-CE of 0.45 Å, 0.58 Å, 0.49 Å, 0.62 Å and 0.51 Å, respectively. Some of these regions are at or near the dimer interface or in close proximity to the substrate-binding site, and hence the differences may be significant for the altered substrate specificity. The largest differences between Cαs of the two structures, however, occur in the region 269–279, with rmsds in the range 0.58–0.94 Å. This helical region (α7) is to the exterior of the subunit and has relatively higher atomic displacement parameters than the bulk of the protein.

A view of the active sites of CRL and lino-CE superimposed in Fig. 7b shows differences in amino acid side chains that may be responsible for their different substrate specificities. Side chains that are almost within van der Waals distance of the cholesterol part of the substrate are Ile127, Ile132, and Ala450 in lino-CE (red in Fig. 7b). Substitutions of Val127 and Thr132 in CRL by Ile127 and Ile132, respectively, in CE enhance hydrophobicity and constrict the cleft. The Cδ-atoms of both these side chains, absent in the CRL structure, make van der Waals contacts with the cholesteryl moiety of the substrate. The substitution Ser450 in CRL by Ala450 in CE is perhaps most significant in the sense that such a change accommodates the strongly hydrophobic cholesteryl moiety in the hydrophobic cleft, by preventing a direct steric conflict with the polar hydroxyl of the serine residue. These differences may be responsible for the change in enzyme specificity from triacylglycerol esters to cholesteryl esters. Substitutions Phe344→Val and Phe455→Trp may also have some effect on substrate selectivity.

There are fewer sequence differences between CE and CRL at the fatty acid-end of the substrate-binding pocket. All the side chains that are in van der Waals contact with the fatty-acid chain are conserved, except for Ala414, which is a glycine in CRL. The Cα-atom of Ala414 is close to the van der Waals contact distance from carbon atoms C11 and C14 of the linoleate chain and may determine the selectivity of CE for fatty-acid chains having *cis* bonds at position 9 or 12, or both (Figs 6,7b).

Other substitutions of possible significance are those at the interfacial openings, or lining the hydrophobic cavity between the two subunits. The substitutions of Tyr69, Pro74, Ala76 and Glu88 in CRL for Phe69, Gly74,

Thr76 and Gln88, respectively, in CE are noteworthy. Substitution of Glu88 for glutamine removes a negative charge from the dimer interface. These changes, along with substitutions such as Ser91→Leu, Ser93→Gln, Thr130→Pro, Ser131→Thr, Thr132→Ile, Ser153→Ala, Glu287→Leu and Thr347→Leu in CE, around the flap region and the central hydrophobic cavity, produce enhanced hydrophobicity in the binding pocket. Hence, binding of more hydrophobic substrates such as cholesteryl esters is favored relative to triacyl glycerols. Finally, substitutions such as Thr5→Lys, Val313→Lys, Ala321→Lys and Glu491→Lys in CE are in the loops exterior to the dimer and probably lead to an increase in the isoelectric point. Important sequence differences between CE and CRL, and their possible significance are summarized in Table 1.

Table 1. Amino acid residue differences between CRL and CE that may influence substrate specificity and/or dimer formation.

Substitutions	Location	Comments
V127I*, T130P, S131T, T132I*, S153A, S159A, M216L, A249P, F344V, G414A*, L417H, V444I, L445M, S450A*, F455W	At or near the catalytic triad/ substrate-binding site.	Except for two residues (F344L and L417H) all these changes enhance the hydrophobicity of the catalytic pocket.
Y69F, P74G, A76T, E88Q	In the flap.	Interfacial helix at the opening to the active site.
S91L, S93Q, V127I, T130P, S131T, T132I, S153A, E287L, F344L, T347L	Near the flap, pointing towards the catalytic cavity/flap.	Renders more hydrophobic or has contacts with the substrate. May influence flap conformation.
T5K, V313K, A321K, E325D, K488N, E491K	External loops.	Exposed loops. Increased pI.

*These residues are in van der Waals contact with the substrate.

Biological implications

Cholesterol esterase from *Candida cylindracea* (CE) reversibly hydrolyzes cholesteryl linoleate and oleate, esters of fatty acids that are major components of arterial plaque. The three-dimensional structures of uncomplexed and linoleate-bound CE at 1.9 Å and 2.0 Å, respectively, show a molecular architecture appropriate for an enzyme that uses hydrophobic lipids as substrates, is soluble in an aqueous environment, and is active at the lipid–water interface. The structure of a monomer of CE is nearly identical to the structure of *Candida rugosa* lipase (CRL), a triacylglycerol acyl hydrolase with which it shares 89% sequence identity. The 'flap' that is capable of covering the hydrophobic active-site cavity is open in both CE structures and resembles that observed in CRL. In all three structures, two monomers face each other in a dimeric association, shielding the hydrophobic environments of two catalytic triads from the aqueous medium. The structure determinations reveal that a substrate could gain access to the catalytic triads through openings at the dimer interface, without having to dissociate the dimer. Although the monomers in the uncomplexed and linoleate-bound CE have identical structures, an increase in the interfacial openings was observed in the linoleate-bound structure, which may indicate a structural change upon substrate entry.

The structure of the complex with cholesteryl linoleate reveals the nature of the lipid–protein interaction and evidence of how a fatty-acid chain might enter the protein. The positioning of the cholesterol moiety is equivocal owing to its weak electron density, suggesting at least partial hydrolysis of the substrate and/or dynamical disorder of the cholesteryl end of the molecule. This end of the substrate could, however, be accommodated in the interfacial hydrophobic cavity when modeled. Of the 55 side chains that differ between CE and the homologous triacylglycerol lipase CRL, 23 are found in and around the active site and at the dimer interface. Some of these side chains are within van der Waals contact distances from the cholesteryl and linoleate ends of the substrate and may influence substrate specificity. Furthermore, the combined effect of these sequence changes is to enhance hydrophobicity in the active site gorge and at the dimer interface. The apparent stability of the dimer suggests that it could be a functionally active form of the enzyme.

Materials and methods

Crystallization and data collection

The commercially available enzyme was further purified and three crystal forms were grown as previously described [17,18]. Lino-CE crystals were grown from the monomer fraction in the presence of cholesteryl linoleate, as described [18]. Apo-CE crystals were obtained from the dimeric fraction of the purified enzyme, without any addition of the substrate [17,18]. The two triclinic crystal forms had one dimer in the unit cell, diffracted well, and were the most suitable for X-ray study. The monoclinic crystal form having four monomers in the asymmetric unit was not pursued. The diffraction data for apo-CE (a=58.57 Å, b=88.69 Å, c=58.59 Å, α=93.29°, β=113.82°, γ=95.98°) were collected at the UCSD Mark III multiwire detector [135378 total reflections; 68305 unique to 1.9 Å; $R_{merge}(F^2)$=9.9%)]. The data collection for the lino-CE (a=58.79 Å, b=90.92 Å, c=58.79 Å, α=93.39°, β=105.74°, γ=97.22°) was performed on an RAXIS IIc image plate detector and processed with the program package from Rigaku (version 3.40) [119382 total reflections; 62706 unique to 2.0 Å; $R_{merge}(F^2)$=6.9%]. The data sets were 92% and 88% complete, respectively, to 2.20 Å, and were 75% and 78% complete, respectively, up to the maximum resolution. In the highest 0.1 Å shell, the data sets were 25% and 36% complete, respectively.

Structure solution and refinement

The structure was determined by the molecular replacement method using the AChE structure [3] as the search model. Calculation of self- and cross-rotation functions for apo-CE

were performed using a combination of X-PLOR [26] and MERLOT [27] routines. The self-rotation function produced a 12σ peak (at position ϕ=–10°, ψ=57°, χ=180° in spherical polar coordinates, with unit cell axis a* along the X-axis in the horizontal plane and c-axis along Z in the vertical direction), while the two cross-rotation peaks had heights of ~6σ (more than 1σ over the background peaks) in MERLOT calculations. A translation problem between two subunits was solved by X-PLOR as well as by a routine written by Dr David Langs [28]. The model was built on a Silicon Graphics Iris Indigo workstation using the program CHAIN, which is a modified version of the program FRODO [29]. This was then subjected to many cycles of X-PLOR simulated-annealing refinement, followed by manual rebuilding. The lino-CE crystal structure was determined starting with partially refined coordinates of apo-CE, followed by similar refinement and rebuilding cycles. Nearly all of the 1068 side chains for both structures could be identified and modeled in the resulting electron-density maps, including the ones at the N and C termini. A large number of isolated, well-ordered peaks were identified and modeled as solvent molecules. Additional electron density attached to asparagine side chains at positions 351 and 314 (both with Asn-X-Thr sequence) were modeled as three *N*-acetylglucosamine molecules, in each monomer.

A model of cholesteryl linoleate was built into the final ($3F_{obs}–2F_{calc}$) and ($F_{obs}–F_{calc}$) maps of lino-CE. The linoleate part of the unbiased density was well defined while the density for the cholesterol part was broken and weak. The cholesterol part of the substrate was therefore modeled in the interfacial space of the dimer, consistent with the fitting of the fatty-acid chain in the well-defined density. The density for the substrate was weaker, in general, in the B subunit than in A. In subsequent cycles of X-PLOR, the positional parameters, isotropic temperature factors, and group occupancy factors of linoleate molecules in both subunits were refined. First, the B-factors for the atoms of the linoleate molecules were fixed at about the average B-values of the neighboring 254 atoms of the linoleate chains in both subunits. The positional parameters of the atoms and two group occupancy parameters were refined, which quickly converged to the values of 0.60 and 0.46 for A- and B-subunits, respectively. In the final cycle of the refinement, occupancy parameters of the two substrate molecules were fixed at these values while the positional and individual B-factors of the substrate atoms were refined. At convergence, the average B-factors of 19 atoms of the linoleate chain were 22.7 Å^2 and 22.0 Å^2 in A- and B-subunits, respectively, while the B-values of 254 neighboring atoms were 16.6 Å^2 and 16.7 Å^2. The average B-values of 28 atoms in the cholesterol molecules were 42.0 Å^2 and 39.1 Å^2 in A- and B-subunits, respectively. The occupancy and thermal parameters of the substrate atoms in the two subunits obtained by this method make physical sense in terms of the observed electron densities at the substrate-binding sites, and also in comparison with the electron densities of neighboring protein atoms. The final difference map in this region exhibited a peak of height of about one-quarter of the average density of the linoleate chain, at the C4-C5 position. A summary of numbers generated from the last cycle of refinement is given in Table 2. Although the limiting resolution of 1.9 Å of the apo-CE structure was higher than that of lino-CE, the data from the lino-CE crystal were intrinsically better and the structure was better determined, as the refinement statistics demonstrate.

Table 2. Refinement statistics.

	Apo-CE	Lino-CE
Protein atoms in the model (2×534 residues)	8074	8074
Solvent atoms:		
water oxygens	336	478
phosphate ions	0	2
Sugar atoms (3 NAGs per monomer)	84	84
Cholesteryl linoleate atoms	–	94
Resolution range	8.00–1.90 Å	8.00–2.00 Å
No. of unique data used ($F^2 > \sigma_F 2$)	49 320	54 351
R-factor	0.168	0.149
No. of unique data (F>0)	63 856	61 540
R-factor for all data F>0	0.201	0.165
Average B-factor	9.7 Å^2	20.5 Å^2
Rmsds:		
bond length	0.011 Å	0.010 Å
bond angle	1.71°	1.56°
dihedral angle	24.3°	24.1°
improper angle	1.53°	1.39°
Ramachandran plot statistics		
Residues in most favored region[a]	85.1%	86.0%
Residues in disallowed regions	Ile18, Ser209 & Ile444 but have good density	Same as in apo-CE
Rmsd of main-chain atoms from non-crystallographic two-fold axis	0.21 Å	0.15 Å

[a]Out of 892 non-glycine and non-proline residues.

Quality of the model

Ramachandran plots [30] for the final refined structures of CE in apo-CE and lino-CE crystals, drawn by the PROCHECK routine [31] showed that Ile18, Ser209 and Ile444 in both subunits in both structures were in disallowed conformations, despite having good electron densities. Final density maps were in excellent agreement with the chemical sequence. Of the 534 side chains, 520 could be located in the electron-density maps. The average isotropic B-factors for protein atoms were 10 Å^2 and 21 Å^2 for apo-CE and lino-CE structures, respectively. The isotropic temperature factor plots for the backbone atoms for both structures exhibited peaks at around residues 80, 275, and 490, with highest B-values of 30 Å^2 and 45 Å^2, respectively, for apo-CE and lino-CE structures. The average temperature factor for the apo-CE atoms was nearly half of that of the lino-CE. This may indicate that apo-CE is a more tightly packed structure than lino-CE, and hence has a lower atomic displacement. However, since two data sets were collected at two area detectors using completely different technologies, the B-factors may not be strictly comparable. Luzzati plots [32] for the two structures were drawn to estimate positional errors. The estimated random positional errors from these plots are 0.20 Å and 0.18 Å for apo-CE and lino-CE structures, respectively.

The coordinates are being deposited with the Brookhaven protein data bank (tracking number T5971).

Acknowledgements: This work is funded by NIH grant No. DK26546 and by the Mae Stone Goode Trust Fund. We thank Drs JF Griffin and GD Smith for critically reading the manuscript and suggesting improvements. Thanks are also due to Prof. N-hu Xuong and his staff at the University of California at San Diego for help in data collection. Melda Tugac and Gloria Del Bel of the Graphics Department are thanked for their excellent work with the figure reproduction. Sandra Finken is thanked for the preparation of the manuscript.

References

1. Small, D.M. (1989). Progression and regression of atherosclerotic lesions. Insights from lipid physical biochemistry. *Arteriosclerosis* **8**, 103–129.
2. Grochulski, P., *et al.*, & Cygler, M. (1993). Insights into interfacial activation from an open structure of *Candida rugosa* lipase. *J. Biol. Chem.* **268**, 12843–12847.
3. Sussman, J.S., *et al.*, & Silman, I. (1991). Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science* **253**, 872–879.
4. Schrag, J.D., Li, Y., Wu, S. & Cygler, M. (1991). Ser-His-Glu triad forms the catalytic site of the lipase from *Geotrichum candidum*. *Nature* **351**, 761–764.
5. Uwajima, T. & Terada, O. (1975). Purification and properties of extracellular cholesterol ester hydrolase of *Pseudomonas fluorscens*. *Agric. Biol. Chem.* **39**, 1511–1512.
6. Cygler, M., *et al.*, & Doctor, B.P. (1993). Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases and related proteins. *Protein Sci.* **2**, 366–382.
7. Ollis, D.L., *et al.*, & Goldman, A. (1992). The α/β hydrolase fold. *Protein Eng.* **5**, 197–211.
8. Hide, W.A., Chan, L. & Li, W.-H. (1992). Structure and evolution of the lipase superfamily. *J. Lipid Res.* **33**, 167–178.
9. Brady, L., *et al.*, & Menge, U. (1990). A serine protease triad froms the catalytic centre of a triacylglycerol lipase. *Nature* **343**, 767–770.
10. Derewenda, Z.S., Derewenda, U. & Dodson, G. (1992). The crystal and molecular structure of the *Rhizomucor miehei* triacylglyceride lipase at 1.9 Å resolution. *J. Mol. Biol.* **351**, 491–494.
11. Winkler, F.K., D'Arcy, A. & Hunziker, W. (1990). Structure of human pancreatic lipase. *Nature* **343**, 771–774.
12. van Tilbeurgh, H., Sarda, L., Verger, R. & Cambillau, C. (1992). Structure of the pancreatic lipase–procolipase complex. *Nature* **359**, 159–162.
13. Martinez, C, De Geus, P., Lauwereys, M., Matthyssens, G. & Cambillau, C. (1992). *Fusarium solani* cutinase is a lypolytic enzyme with a catalytic serine accessible to solvent. *Nature* **356**, 615–618.
14. Uppenberg, J., Hansen, M.T., Patkar, S. & Jones, T.A. (1994). The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida antarctica*. *Structure* **2**, 293–308.
15. Brzozowski, A.M., *et al.*, & Thim, L. (1991). A model for interfacial activation in lipases from the structure of a fungal lipase–inhibitor complex. *Nature* **351**, 491–494.
16. van Tilbeurgh, H., *et al.*, & Cambillau, C. (1993). Interfacial activation of the lipase–procolipase complex by mixed micelles revealed by X-ray crystallography. *Nature* **362**, 814–820.
17. Ghosh, D., Erman, M. & Duax, W.L. (1991). Crystallization and preliminary diffraction analysis of cholesterol esterase from *Candida cylindracea*. *J. Steroid Biochem. Mol. Biol.* **38**, 663–665.
18. Kaiser, R., Erman, M., Duax, W.L. Ghosh, D. & Jörnvall, H. (1994). Monomeric and dimeric forms of cholesterol esterase from *Candida cylindracea*. Structure, identity in peptide patterns and additional microheterogeneity. *FEBS Lett.* **337**, 123–127.
19. Labow, R.S., Adams, K.A.H. & Lynn, K.R. (1983). Porcine cholesterol esterase, a multiform enzyme. *Biochim. Biophys. Acta* **749**, 32–41.
20. Rudd, E.A., Mizuno, N.K. & Brockman, H.L. (1987). Isolation of two forms of cholesterol esterase from porcine pancreas. *Biochim. Biophys. Acta* **918**, 106–114.
21. Liao, D.-I., Breddam, K., Sweet, R.M., Bullock, T. & Remington, S.J. (1992). Refined atomic model of wheat serine carboxypeptidase II at 2.2 Å resolution. *Biochemistry* **31**, 9796–9812.
22. Janin, J. & Chothia, C. (1990). The structure of protein–protein recognition site. *J. Biol. Chem.* **265**, 16027–16030.
23. Derewenda, U., Brzozowski, A.M., Lawson, D.M. & Derewenda, Z.S. (1992). Catalysis at the interface: the anatomy of a conformational change in a triglyceride lipase. *Biochemistry* **31**, 1532–1541.
24. Grochulski, P., *et al.*, & Cygler, M. (1994). Analogs of reaction intermediates identify a unique substrate binding site in *C. rugosa* lipase. *Biochemistry* **33**, 3494–3500.
25. Quinn, D.M. (1987). Acetylcholinesterase: enzyme structure, reaction dynamics and virtual transition status. *Chem. Rev.* **87**, 955–979.
26. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R-factor refinement by molecular dynamics. *Science* **235**, 458–460.
27. Fitzgerald, P.M.D. (1988). MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement. *J. Appl. Crystallogr.* **21**, 273–278.
28. Langs, D.A. (1975). Translation vector functions based on deconvolution of the Patterson function provided by transform methods. *Acta Crystallogr. A* **31**, 543–550.
29. Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272.
30. Ramachandran, G.N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–437.
31. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
32. Luzzati, V. (1952). Traitment statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr.* **5**, 802–810.
33. Evans, S.V. (1993). SETOR: hardware lighted three-dimensional solid model representation of macromolecules. *J. Mol. Graphics* **11**, 134–138.

Received: **3 Jan 1995**; revisions requested: **25 Jan 1995**; revisions received: **3 Feb 1995**. Accepted: **6 Feb 1995**.